DOCUMENT RESUME

ED 368 210 FL 022 007

AUTHOR Skinner, Robert E.
TITLE LEP Readers: Standardized Testing versus Informal Testing.
PUB DATE [94]
NOTE 16p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Comparative Analysis; *English (Second Language); *Limited English Speaking; *Reading Tests; *Standardized Tests; Test Reliability; *Test Use; Test Validity

ABSTRACT

The merits and disadvantages of standardized and informal reading tests for limited English proficient readers are discussed. A growing reliance on standardized ("formal") tests due to their ease of administration and scoring is criticized because the tests are seen as: inadequate for describing students at high and low ends of the scale; not readily interpreted for this population; requiring too much independent work or background knowledge; culturally biased; and generally valid or reliable in some aspects but not in others. "Informal" testing is found advantageous because: it can occur over a variety of contexts, skills, and focal points; provides learning and low emotional and academic risk; and uses varied measurement techniques, selected for their appropriateness to the situation and population. Examples and supporting data are drawn from recent research. It is concluded that formal testing has multiple limitations and should be used with great caution. A trend toward individualized testing, as through student portfolios, is seen as positive, while comparison among individual readers is viewed as counterproductive. In addition, it is proposed, reading assessment should be a learning opportunity, relevant to the individual's progress, which can only be accomplished with informal, non-standardized testing. (MSE)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED 368 210

Author: Robert E. Skinner

# LEP Readers: Standardized Testing Versus Informal Testing

Test! Test! Test--- ad infinitum! To test or not to test is not the question. Evaluation is a given in the educational setting, but "How does assessment provide the best measure of student performance?" is the billion-dollar query for educators.

No measure can be considered one-hundred percent accurate simply because each individual is unique; however, it appears that the enormity of the task, complicated by the vast numbers of the multi-cultural, multi-ethnic populations of varied socio-economic levels tested, has propelled the test-construction industry into a frenzy of standardized forms. Comparative statistics are compiled from results on norm-referenced measures in order to establish a guide by which testees can become aware of how they compare with each other on particular tests. No criteria exist to compensate for individual differences; therefore, the validity of the comparisons is questionable. Furthermore, criterion-referenced tests are designed to measure the accomplishment of certain objectives; however, the immense variety of curricula at innumerable levels on the same grade level in different schools in different communities precludes the accomplishment of objectives on the same level. Finally, standardized tests may be used to establish quotas, such as "pass" or "fail" or for placement in or out of certain levels. Educational opportunities become more exclusive for students on both ends of the success scale. Students may score poorly on a standardized exam but may perform very well on day-to-day academic tests. It is a discriminatory practice to use a standardized test form as the exclusive measure for evaluation. Thus, the controversy rages---

FL022007



"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Robert E. Skinner
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it
☐ Minor changes have been made to improve reproduction quality
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy



BEST COPY AVAILABLE

standardized measurement versus authentic assessment or informal testing.

## Standardized Testing

The pragmatist in the formal assessment arena sees standardized testing as the only option for an objective evaluation of the reader. Undoubtedly, this advantage allows for the same test to be administered to enormous numbers according to standard guidelines. There is no place for various interpretations of answers. Each question has only one specified answer. When tests are scored, different scorers always come up with the same score. As a result, measurement results in a certain number of correct responses and a certain number of incorrect responses. "The psychometric description of norm-referenced testing takes classical test theory as their starting-point, which is based on a normal distribution of test scores; i.e. a few high scores, a few low scores, and most scores around the mean.... The ideal item in a norm-referenced test is one that is scored by 50% of the learners. An item which 90% of the learners get right is a bad item in this view, because it hardly differentiates between good and poor learners. If, however, one is interested in the question whether or not a learner has mastered something, such an item might very well be a good item, and the same holds for items which are too difficult. In norm-referenced test items which are too easy or too difficult are excluded; a result of this procedure may be that parts of the domain to be tested are excluded from the testing."(van Els, Bongaerts, Extra, van Os, van-Dieten, 1986) Advantages of the norm-referenced test allow the testee to know his rank on the test as compared to all other testees. The criterion-referenced tests provide a profile of the testee's reading skills. The testee demonstrates what he or she can or cannot do (Baker, 1993); thus, he or she may pass or fail based on his or her test score. All testees who can do the tasks

required, pass; all who cannot, fail.

There are studies, without number, listing the barrage of test batteries bombarding the student arena. Reading of the literature reveals scores of tests and scores of comparisons. Norms are cited and judgment calls are made. A survey of the literature shows, however, that more data is available on norm-referenced tests than on criterion-referenced tests. Perhaps one of the reasons for this occurrence is that difficulty in establishing what is valid or not rises from another difficulty--- "What does it mean to have mastered an item?" (van Els,1986) On the other hand, it is very clear that formal testing has a historical basis, considered time-tested and traditional. Consequently, it is hypothesized that teacher judgments and results on standardized exams would be relatively commensurate if teachers were not aware of the English reading ability of LEP students and that teachers' judgments would be less consistent if the teachers were aware of the students' reading abilities in English. A study concluded that the first hypothesis was true after students took the (LAS) or Language Assessment Scales. The study indicated that a teacher's judgment can be subjectively influenced to the point that objective assessment is lost (Sims, Levine, Eaton, Cabello, 1989). Furthermore, it is assumed by proponents of formal testing that test success is a cognitive measure rather than an affective or enactive one, but that the cognitive will transfer to the enactive. This attitude is described as being typically present in the United States (Finlay and Harrison, 1992). Standardized tests receive a rating and are normally given to large samples. When they are representative of the population for which they are designed, they receive a rating of "Good". For example, the SDRT and WRMT-R were rated "Good" in this domain(Lewandowski and Martens, 1990). Various other categories, such as content, reliability, validity, theoretical base, utility, and scores are scrutinized for ratings. The relatively large number of standardized tests and test manuals, which provide directions for taking the tests,

and the relative ease of administration and scoring adds to the advantages for using them. Obviously, their greatest advantage is their availability.

## Informal Testing

Nonetheless, informal testing of reading clearly specifies the limitations of standardized testing. In fact, it is suggested that "One may think of the assessment situation as a continuum with standardized objective testing at one end and informal subjective assesment at the other." (Finlay and Harrison, 1992) At this end of the continuum a concensus in the literature can be found just as outspoken as that of the proponents of formal testing. Reasons for denigrating the universal and collective usage of standardized exams are many. Perhaps the most effective manner to approach this subject is to cite the most blatant arguments by the opponents. Indeed, it is most noteworthy to recognize that standardized tests discriminate. That is, "Group tests are generally poor for describing children at low or high ends of the distribution (Bracken, 1988) because they typically contain few items that are very easy or difficult; thus, a limited number of test items is an inadequate measure of any skill (Lewandowski and Marten,1990). Standardized scores are considered objective and without error; however, human error can also occur in interpretation of test scores. Should a percentile rank discourage or encourage? (MacGinitie, 1993) Moreover, the Education Commission of the States notes that "(1) Reading performance is generally lower for OL (other languages) students than for EL (English language) students; (2) Low socioeconomic status shows a high correlation with low achievement; (3) Language dominance seems to have different effects for people of different ethnic and cultural backgrounds; (4) OL students are not evenly distributed throughout the country in various schools, or within population groups; and (5) while many OL

students are Hispanic, many belong to other ethnic and cultural groups (National Assessment of Educational Progress, 1982). What these findings indicate is that a standardized exam does not take into consideration the individual differences and diversity of the testees. In addition, students tested in their mother tongue always score higher on reading tests than when they are tested in another language; consequently, if norm-referenced tests compare bilinguals to monolinguals, the measurement is invalid (Baker, 1993). Again, it must be emphasized that reading difficulty can be due to inadequate development in the target language (King and Quigley, 1985). Besides this, there is evidence that reading differences, not considered by standardized tests, may be the result of differences in the "Home support-system," where English may not be spoken (Baratz-Snowden and Swan, 1987). In addition, standardized testing may ask kinds of questions which are misunderstood. If the questions were asked differently, the LEP student may read with understanding (Langer, Judith, and others,1990). It is also interesting to note that test results indicate that LEP students comprehend better when they have been read to, rather than when they read independently (Zuniga-Hill and Parsons, 1991). Obviously, many standardized exams require totally independent work. Likewise, if LEP students recognize cognates on multiple-choice tests, their comprehension increases (Nagy and other, 1992), but many cognates may not be present. If students also lack the background knowledge, they have greater reading difficulties on standardized tests (Pritchard, 1990). Other reasons why standardized testing is rejected are cited by Finlay and Harrison (1992): the testee may be very anxious, may not have well-developed test-taking strategies, may learn how to guess well (a poor strategy), may do well on the test but not be able to function in real-life situations, may be tested over what is "in theory" taught but "not actually taught," the test may be designed with isolated questions independent and "autonomous" of situations, and

contrary to the intent of the test designer, there may be several valid interpretations of a text (Finlay and Harrison, 1992). In like manner MacGinitie reveals three biases built into formal testing. These biases, evidenced in experiments, show that negative impressions are formed when certain words are used. That is to say, certain words have negative connotations like "cold" when it is substituted for "warm" when describing people; furthermore, a category bias exists when people are associated with certain occupations. For example, the idea of a "waitress" or a "librarian" bring biases to the understanding of those words especially when different interpretations are given to the meaning of those occupations. The bias will determine how a test question is answered. Thirdly, there is a"confirmation" bias concerning the influence of the different beliefs held by different individuals. One's beliefs may blind him from seeing alternative answers. He can see only what he believes (MacGinitie, 1993). Another bias, the "content" bias is discussed by Lewandowski and Martens. They say that the content of the test may not be consistent with the curriculum (Lewandowski and Martens, 1990). Also, standardized tests may be apprised as good in some areas and still be given even though they are assessed fair or poor in other areas; thus, even the instrument is called into question by those who support it. For instance, the WRMT-R and SDRT were apprised "good" in their representation of the population tested; however, WRMT-R received a poor rating and SDRT, a fair rating, for their theoretical bases. The SDRT appears to test for which it is designed, but on the contrary, WRMT-R appears to test something else. Both received good ratings for their score profile, but neither score any higher than fair for their validity (Lewandowski and Martens, 1990). Undoubtedly, even standardized tests, when measured, don't always measure up!

With many of the pros and cons of formal testing

presented, it is clear that much controversy exists concerning the standardized test, but there is also evidence to support informal assessment. "Increasingly, teachers are making clear that they know how to address accountability issues through good documentation of children's actual work." (Perrone, 1993) In the informal arena authentic assessment appears time and time again. Emphasis is on what the student can do rather than on what he cannot do. The advantages of informal assessment are many. Assessment can occur over a variety of contexts, skills, and focal points, which may indicate that a student has weaknesses but strengths in other areas (Wolf, 1993) Standardized exams are not aware of those being tested, but teachers are informed and knowledgeable of their students. Wolf says it best, "...assessment is informed, rather than informal, when it is carried out by knowledgeable teachers who draw on a variety of strategies to carefully observe and document their students' performances across diverse contexts and over time as students are engaged in authentic learning tasks (Wolf, 1993). Informal assessment also provides "...positive learning and low emotional and academic risk...." where real world tasks are measured and the process used by students to arrive at answers is demonstrated and recorded (Peers, 1993). Michele Peers developed a task to utilize performance-based assessment as a "diagnostic tool" in her classroom. Her intent was to integrate reading and writing in an informal assessment task that normally would be tested on a formal test. Peers discovered that her students had great difficulty with the task because it involved higher-order thinking strategies; however, the informal test made her aware of the "gaps" in her students' learning as well as the strengths of her students upon which she could build in the classroom. Post-evaluation exercises allowed for cooperative peer learning (Peers, 1993). None of this could have been possible within the scope of a standardized test. Again, it must be noted that standardized test designers do not know the students for whom the tests are designed. Nor

do they allow for the conditions best supporting student success. Informal assessment considers the value of a variety of ways of measurement. For instance, cooperative learning groups can be observed sharing strategies which provide the help that bilingual readers need to develop meaning as they read. (Walker, 1989) Assessment can occur while reading occurs in an informal setting. Paul Simmonds (1985) points out that "...the most effective tests involve transfer of information, require students to interpret speech, involve cooperative tasks, require students to process and synthesize information in a variety of ways, and bring in authentic materials." He continues "These tests may provide a more useful profile of a candidate, especially when such questions are set within a clearly expressed and natural framework." Again, his evaluation of assessment insists on anything but a standardized format, and he bases his conclusion on a comparison of tests from examination boards providing exams for ESL and EFL students. An innovative approach of informal assessment is shared by John Harker (1985). According to Harker a valuable assessment tool for reading is the diagnostic tool called retelling. The process involves reading of various materials followed by a retelling of what was read. Comprehension, of course, can be very quickly measured in this exercise. The Hispanic student may have difficulty in comprehending an English reading passage, but Robert Goslin (1978) indicates that the "...second strongest indicator of subject matter achievement was reading ability when assessed in Spanish." Again, the standardized test ignores the literacy strengths of the bilingual student in his mother tongue, but the informal measurement can consider his mother tongue as well as his cultural background. It is obvious that no examination can measure everything that results in a student's success, but there is no doubt about how a formal test tends to homogenize everyone when, in reality, test-taking society is heterogenous.

## POSITION

In consideration of the foregoing conclusions in support of informal assessment, enough evidence indicates that formal testing is one option with various limitations and should be used with great caution. The greatest disservice a formal test can have is to whittle away an individual's self-esteem each time he faces another formal measurement. Is it too hyperbolic to say "By the time the three days of real testing are over, weeks, sometimes months, have passed. Time for real books has been sacrificed for time spent reading isolated paragraphs and answering multiple-choice questions." (Perrone, 1993)? It is widely known in the United States that governmental agencies require accountability because accountability is tied to dollars. For funds to be doled out, there must be some measure that can be used to justify the expenses. Formal testing seems to fit the bill since results are easily obtained and can be quickly compiled for comparative data whether that data is valid or not. Authentic assessment requires time and real-life performance that is neither readily accessible nor easily comparable. Neither informal assessment time nor real-life performance meet the arbitrary time-table tied to government dollars. Somewhere along the way, the individual has been lost. He is drowning in a deluge unjustly assessing his real abilities. The clarion call is for educational and governmental agencies to remove their blinders in order to take a clear look at the differences between mass testing on standardized exams and actual individual performance. Should one take a look at assessment outside of the United States, standardized assessment is not always regarded so highly. In England, standardized assessment is criticized in Adult Basic Education, while customized literacy tests, which measure functional ability, can be found (Finlay

and Harrison, 1992). The English program is not without weaknesses, but it serves to remind the academic community that reading skills are functional skills measurable on the basis of what an individual can do with those reading abilities. It must be realized that there is validity in a variety of methods of assessment, which "... offer an array of windows on student learning " (Wolf, 1993). In 1983 "A Nation at Risk" was published. It's powerful influence and a tremendous amount of pressure since the seventies make the academic community keenly aware that standardized testing has ruled the day (Perrone, 1993), but a movement towards individual assessment is once again blowing in the wind, and it is being received very favorably in various academic circles--- the public schools and in higher education. "Student portfolios are an appropriate assessment device for any age group...." (Stahle and Mitchell, 1993) A portfolio can provide "...a complete picture of a student's literacy abilities." (MacGinitie,1993) Although the portfolio contains writing samples, rather than reading samples, and there are questions about how to assess what is placed in a portfolio, specific reading success records about specific strengths and weaknesses,as well as spontaneous moments indicating student ability can be included in the portfolio. Cathy Grace says, "The portfolio is a record of the child's process of learning: what the child has learned and how she has gone about learning; how she thinks, questions, analyzes, synthesizes, produces, creates, and how she interacts/ intellectually, emotionally and socially/with others." Grace continues to relate that evaluation is based upon comparing former work accomplished to current work completed.The norm would be the individual's progress, not a comparison between the student and other students but a measure of a "student's progress over time." (Grace, 1993)

What is the real issue of reading assessment? Obviously, it all rests on this fulcrum: Can a student read or not? It is admitted "A test score hardly ever exactly represents the

true performance of the testee; it may underestimate or overestimate his performance." (van Els, 1986) Then, what really should be concluded? Not any single test can measure the reading abilities of all testees. If a test can not be used in all situations for all testees; then, there must be alternative valid and reliable means of assessment; otherwise, unjust academic calls may be made. Never should a student be considered more successful than another because he has developed a test-taking behavior for successful test-taking. Moreover, a reading test should be very cautious when it provides a grade equivalent score, since such information means little or nothing except that the test itself does not consider that reading abilities are in flux between the elementary and high school years. Furthermore, when statements from the academic community like the following are made, educators need to ask, "Why?" "All testing of young children in preschool and grades K-2 and the practise of testing every child in the later elementary years should cease. To continue such testing in the face of so much evidence of its deleterious effects is the height of irresponsibility." (Perrone, 1993) Too often the recommendation to the test administrator is to know the instrument and what it tests. This recommendation presupposes that the prescription for assessment is another standardized test for every academic ailment. Do standardized tests create better readers? If reading is to be authentically assessed, the teacher must be allowed "...the freedom to situate or contextualize assessment." (Garcia and Pearson, 1991) Of course, the teacher would then have to increase his "...knowledge base about language, culture, and literacy." (Garcia and Pearson, 1991)

Assessment should not exist to screen someone out but to let someone in. Reading assessment should be a learning experience, an open-door, and an opportunity. Never should it be a programmed comparison. It should be relevant to the individual and accurate about his performance. Should formal tests "... contain words that children likely have never seen

and certainly don't use." (Perrone, 1993)? How relevant is that kind of test, and what does it prove? Never is the question: "How many did he or she answer incorrectly?" Rather, the question is "Can he or she read? or "What can he or she do because he or she can read?" Test! Test! Test! Yes, but the question is "How do we best measure an individual's reading ability?" Will an LEP student be assessed inaccurately by a monolithic standardized examination? Evidence subtantiates this possibility. The decision, then, is for informal over formal assessment.

BIBLIOGRAPHY

Baker, Colin, Foundations of Bilingual Education and Bilingualism. Multilingual Matters, Ltd. Clarendon. Philadelphia, 1993.

Baratz-Snowden and Richard Duran. "The Educational Progress of Language Minority Students." Findings from The 1983-84 NAEP Reading Survey National Assessment of Educational Progress. Princeton, New Jersey, 1987.

Finlay, Ann and Colin Harrison. "Measuring 'Success' inReading in Adult Basic Education: A United Kingdom Perspective." Journal of Reading, vol. 36, no.3 (Nov. 1992), pp. 164-172.

Garcia, Georgia and David Pearson. "Literacy Assessment in a Diverse Society." Center for the Study of Reading. Champaign, Illinois, 1991.

Goslin, Robert Dilwyn. "Language and Reading Factors as Indicators of Achievement in Science and Social Studies for Students in a Bilingual Education Program." University Microfilms International Dissertation Copies, 1978.

Grace, Cathy. "The Portfolio and Its Use: Developmentally Appropriate Assessment of Young Children (from the ERIC Digest) Texas Association of Administrators and Supervisors of Programs for Young Children, vol1, no.1 (Fall, 1993), pp. 8-9.

Harker, W. John, ed. Classroom Strategies for Secondary Reading. Second edition. International Reading Association, 1985.

King, Cynthia and Stephen Quigley. Reading and Deafness. College Hill Press. California, 1985.

Langer, Judith. "Meaning Construction in School Literacy Task: Study of Bilingual Students." American Educational Research Journal, (Fall, 1990).

Lewandowski, Lawrence J. and Brian K. Martens. "Selecting and Evaluating Standardized Reading Tests." Journal of Reading, vol. 33, (Feb., 1990), pp. 384-388.

MacGinitie, Walter H. "Some Limits of Assessment." Journal of Reading, 36:7, (April, 1993), pp. 556-559.

McAuliffe, Sheila. "A Study of the Difference Between Instructional Practice and Test Preparation." Journal of Reading, 36:7, (April, 1993), pp. 524-530.

Nagy, William E. and others. "Cross-Language Transfer of Lexical Knowledge: Bilingual Students Use of Cognates." Center for the Study of Reading. Champaign, Illinois, 1992.

Peers, Michele G. "A Teacher/Researcher's Experience with Performance-Based Assessment of a Diagnostic Tool." Journal of Reading, 36:7 (April, 1993), pp. 544-548.

Perrone, Vito. "On Standardized Testing." ( from the ERIC Digest) Texas Association of Administrators and Supervisors of Programs for Young Children,vol 1, no. 1 (Fall, 1993), pp. 6 - 7.

Pritchard, Robert. "The Effects of Cultural Schemata on Reading Processing Strategies." Reading Research Quarterly, vol. 25, no. 4, (Fall, 1990), pp. 273-295.

Simmonds, Paul. "Survey of English Language Examinations." ELT Journal, vol. 39. no.1, (January, 1985), pp. 33-42.

Sims, Antonia Lopez, Martin G. Levine, Arlinda Eaton, and Beverly Cabello. "Are Teacher Judgments of Elementary LEP Students' English Oral Proficiency Affected by Their Judgments of Students' English Reading Ability?" Journal of Educational Issues of Language Minority Students, vol.4, (Spring, 1989), pp. 73-83.

Stahle, Debra L. and Judith P. Mitchell. "Portfolio Assessment in College methods courses: Practicing What We Preach." Journal of Reading, 36:7 (April, 1993), pp. 538-542.

Van Els, Theo, Theo Bongaerts, Guus Extra, Charles van Os, Ann-Mieke, Janssen-van Dieten. Applied Linguistics and the Learning and Teaching of Foreign Languages. Edward Arnold Publishers Ltd. Baltimore, Maryland, 1986.

Walker, Barbara, J. "A reading Strategies Program for Native American Students." Proceeedings of the Annual International Native American Language Issues (Billings, Montana), June 8 - 9 , 1989, pp. 121-132.

Wolf, Kenneth P. "From Informal to Informal Assessment: Recognizing the Role of the Classroom Teacher." Journal of Reading, 36:7, (April, 1993), pp. 518-522.

Zuniga, Hill Carmen and Diana Parsons. "The Effect of Listening on Text Comprehension for Limited English Proficient

Students." National Clearinghouse for Bilingual Education, Document 57 (BEO18270), 1991, 12 p.